Self-consistent solution to the Flory-Huggins model and applications in protein liquid-liquid phase separation

Daoyuan Qian,[1] Thomas C. T. Michaels,[2,3,1,†] and Tuomas P. J. Knowles[1,4,†]

[1] *Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK*

[2] *Department of Physics and Astronomy, Institute for the Physics of Living Systems, University College London, London, UK*

[3] *Laboratory for Molecular Cell Biology, University College London, London, UK*

[4] *Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Ave, Cambridge, CB3 0HE, UK*

A self-consistent analytical solution for the binodal of the 2-component Flory-Huggins model is derived. We show that this form extends the validity of the Ginzburg-Landau expansion away from the critical point to cover the whole phase space, from which an exponential scaling law of the dilute phase binodal concentration is derived. This approach allows us to fit experimental protein liquid-liquid phase separation data and extract parameters including the effective chain length and the solute-solvent interaction energy. We demonstrate explicitly the power of this approach by quantifying the interaction energy contributions of individual amino acids. This analytical framework can serve as a new way to decode the protein sequence grammar for liquid-liquid phase separation.

I. INTRODUCTION

The formation of membraneless organelles through liquid-liquid phase separation (LLPS) has emerged as an important mechanism used by cells to regulate their internal biochemical environments, and it is also closely related to development of neurodegenerative diseases [1–7]. The Flory-Huggins model [8–10] is a foundational theoretical picture that describes the phenomenology of LLPS, driven by a competition of entropy and interaction energy. Despite the generality of the Flory-Huggins model, analytical solutions describing the binodal line have not been available, and progress has instead been made through numerical methods [10, 11]. Here we propose an analytical self-consistent form for the binodal concentrations and demonstrate the high accuracy comparable to numerical schemes. This can then be used to efficiently fit experimental binodal data and extract physically relevant parameters.

II. THE FLORY-HUGGINS MODEL FOR BINARY MIXTURES

The 2-component Flory-Huggins theory describes mixing of a polymer species of length N with a homogeneous solvent. Denoting the volume fraction of polymers as ϕ, the volume fraction of the solvent is simply $1-\phi$ by volume conservation. Further assuming an effective lattice site contact energy $\chi \equiv \frac{z}{2}\frac{2\epsilon_{\rm ps}-\epsilon_{\rm ss}-\epsilon_{\rm pp}}{k_{\rm B}T}$ between polymer and solvent (in which z is a coordination constant, and $\epsilon_{\rm ps}$, $\epsilon_{\rm ss}$ and $\epsilon_{\rm pp}$ are bare polymer-solvent, solvent-solvent and polymer-polymer contact energies) the free energy density of the Flory-Huggins model is given by [8–11]

$$\frac{f(\phi)}{k_{\rm B}T} = \frac{\phi}{N}\ln\phi + (1-\phi)\ln(1-\phi) + \chi\phi(1-\phi) \qquad (1)$$

where $k_{\rm B}$ is the Boltzmann constant and T the absolute temperature. In the following we set $k_{\rm B}T = 1$ to simplify notations. The first two terms on the right hand side of equation (1) represent the entropic energy of mixing while the third term denotes the effective contact energy.

Spinodal concentrations. The free energy density becomes locally unstable at $f''(\phi) \le 0$, and consequently the spinodal boundary $\phi^{\rm spi}$ is defined at the transition point $f''(\phi^{\rm spi}) = 0$. Solving this we obtain the dense ($\phi_+^{\rm spi}$) and dilute phase ($\phi_-^{\rm spi}$) spinodal concentrations

$$\phi_\pm^{\rm spi} = \left(\frac{1}{2} - \frac{\gamma}{4\chi}\right) \pm \sqrt{\left(\frac{1}{2} - \frac{\gamma}{4\chi}\right)^2 - \frac{1}{2\chi N}}, \qquad (2)$$

with $\gamma \equiv 1 - \frac{1}{N}$, which goes to zero in the symmetric $N = 1$ case. The critical point of LLPS occurs when the dense and dilute phases coincide, corresponding to a critical interaction strength $\chi_{\rm c} = \frac{1}{2}\left(1 + \frac{1}{\sqrt{N}}\right)^2$ and concentration $\phi_{\rm c} = \frac{1}{1+\sqrt{N}}$. Near the critical point with $\delta\chi \equiv \chi - \chi_{\rm c} \approx 0$ the spinodal concentrations have the approximate form

$$\phi_\pm^{\rm spi} \approx \phi_{\rm c} \pm \sqrt{\frac{\delta\chi}{2\chi_{\rm c}^2\sqrt{N}}}. \qquad (3)$$

Note that in the opposite limit of large N or large χ the dilute phase concentration has a power-law scaling

$$\phi_-^{\rm spi} \approx \begin{cases} \frac{1}{(2\chi-1)N} & \text{for } N \gg 1 \\ \frac{1}{2\chi N} & \text{for } \chi \gg 1 \end{cases} \qquad (4)$$

and these are qualitatively different from the exponential scaling of the dilute phase binodal concentrations, as

† t.michaels@ucl.ac.uk
† tpjk2@cam.ac.uk

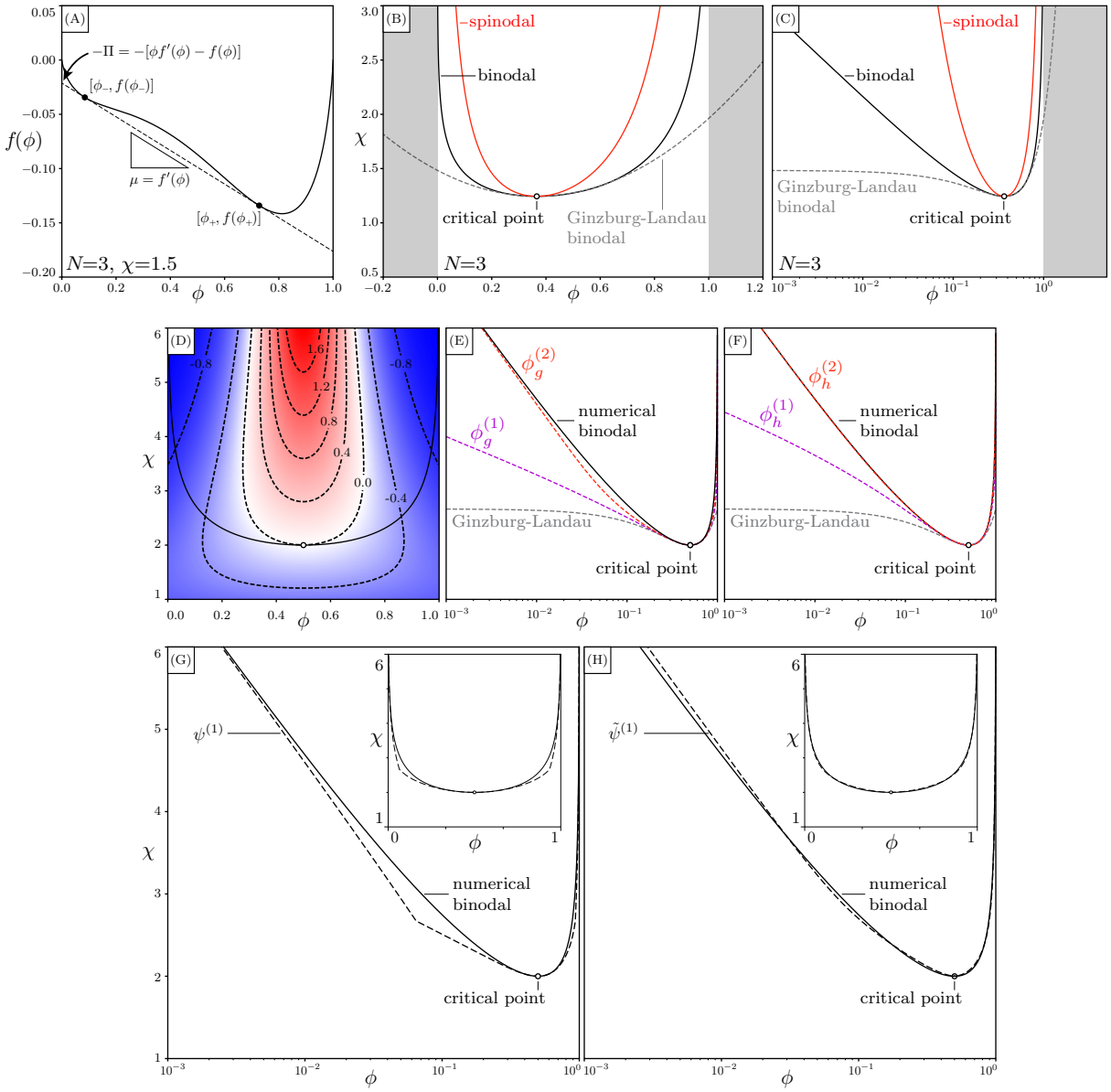

FIG. 1. Flory-Huggins model (A) - (C) and self-consistent solution for the symmetric $N = 1$ case (D) - (H). (A) Common tangent construction at $N = 3$, $\chi = 1.5$ gives the dense and dilute phase concentrations $\phi_{1,2}$. The gradient of the common tangent is the chemical potential $\mu(\phi)$ and the y-intercept is -1 times the osmotic pressure $\Pi(\phi)$. (B), (C) Complete phase diagram of $N = 3$ system in linear and logarithmic ϕ scales. The binodal is calculated numerically [10]. Near the critical point the Ginzburg-Landau binodal approximates the exact binodal well, but at large χ the two quickly diverges and the Ginzburg-Landau solution enters the unphysical range $\phi < 0$ and $\phi > 1$ (grey zones). (D) Plot of $|g'(\phi)| - 1$ in ϕ, χ space. Black solid line is binodal and hollow circle is the critical point. Dashed lines are contours of constant $|g'(\phi)| - 1$. The blue region with $|g'(\phi)| - 1 < 0$ has stable orbits while red regions are unstable. (E), (F) Comparison between the numerical binodal (black solid line) with self-consistent schemes with 0, 1 or 2 iterations (coloured dashed lines). $\phi_g^{(i)}$ is the original map and $\phi_h^{(i)}$ the improved map. The improved map is indistinguishable from the numerical results after 2 iterations. (G), (H) Approximate analytical expressions $\psi^{(1)}$, $\tilde{\psi}^{(1)}$ (black dashed line) give good convergence to the binodal (black solid line). Insets: (G), (H) in linear scale.

will be shown using the self-consistent equations.

Binodal concentrations. The binodal concentrations are found by assuming the existence of two distinct phases characterised by polymer volume fractions ϕ_+, ϕ_-, and compartment volumes V_+, V_-. Equilibrium condition requires minimisation of the total energy $F_{\text{tot}} \equiv V_+ f(\phi_+) + V_- f(\phi_-)$ subject to total volume and mass conservation conditions $V_+ + V_- = V_{\text{tot}}$ and $V_+\phi_+ + V_-\phi_- = V_{\text{tot}}\phi_{\text{tot}}$. Using Lagrange minimi-

sation we identify the chemical potential $\mu(\phi) \equiv f'(\phi)$ and osmotic pressure $\Pi(\phi) \equiv \phi f'(\phi) - f(\phi)$ as Lagrange multipliers that have to hold the same values in the two compartments:

$$\begin{cases} \mu(\phi_+) = \mu(\phi_-) \\ \Pi(\phi_+) = \Pi(\phi_-) \end{cases} . \tag{5}$$

Graphically, in a $[\phi, f(\phi)]$ plot the $\mu(\phi_+) = \mu(\phi_-)$ condition forces the two points to have the same gradient and $\Pi(\phi_+) = \Pi(\phi_-)$ aligns the two tangent lines to have the same y-intercept, so together they are called the common tangent construction (figure 1A). Using the Flory-Huggins expression (1) we have

$$\begin{aligned} \mu(\phi) &= \chi(1-\phi) - \chi\phi + \frac{\ln\phi}{N} - \ln(1-\phi) - \gamma \\ \Pi(\phi) &= -\gamma\phi - \chi\phi^2 - \ln(1-\phi) \end{aligned} \tag{6}$$

and the binodal concentrations can be calculated by solving (5) with the definitions of (6). The objective of this paper is to generate approximate analytical solutions of these. As a first step, an approximate binodal solution near the critical point can be worked out by expanding the free energy around $\phi = \phi_c + \delta\phi$ and $\chi = \chi_c + \delta\chi$ with small $\delta\chi$ and $\delta\phi$. Terms that are constant or linear in $\delta\phi$ drop out of the common tangent construction; terms of order higher than 4 are also truncated. The result is $f(\phi_c + \delta\phi) \approx -\delta\chi\delta\phi^2 + \frac{\sqrt{N}\chi_c^2}{3}\delta\phi^4$ and for $\delta\chi > 0$ this is a simple Ginzburg-Landau second order phase transition. The binodal concentrations near the critical point are

$$\phi_\pm^{\text{bin}} \approx \phi_c \pm \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}. \tag{7}$$

The Ginzburg-landau solution describes the binodal near the critical point (figure 1B, C) and we can extend this solution to cover χ far away from χ_c through a self-consistent approach, summarised as the following. Suppose we need to solve the equation $\eta = \mathcal{A}(\eta)$ with some operator $\mathcal{A}$. Instead of solving it directly, we treat $\mathcal{A}$ as a discrete map and start with a guess solution $\eta^{(0)}$ and apply $\mathcal{A}$ iteratively to generate the orbit $\eta_{\mathcal{A}}^{(i)} \equiv \mathcal{A}^i[\eta^{(0)}]$ with $i = 1, 2, \ldots$. With a suitable form of $\mathcal{A}$ the orbit converges to the fixed point $\lim_{i\to\infty} \eta^{(i)} = \eta^*$, which then solves the equation $\eta^* = \mathcal{A}(\eta^*)$. This is the contractive mapping principle [12, 13]. The self-consistent approach has been previously employed to approximate the protein aggregation kinetics curves [14, 15] and here we show a similar procedure allows efficient and accurate computation of the binodal concentrations too.

III. SELF-CONSISTENT SOLUTION FOR $N = 1$

Self-consistent equation. With unit polymer length $N = 1$ the free energy (1) is symmetrical, and the binodal is given by the condition $f'(\phi) = 0$, leading to $\frac{1-\phi}{\phi} = \exp(-2\chi\phi + \chi)$. Upon rearrangement the binodal equation is

$$\phi = \frac{1}{1 + \exp(-2\phi\chi + \chi)} \equiv g(\phi) \tag{8}$$

and we use $g(\phi)$ to define the 1D map $\phi_g^{(i)} \equiv g[\phi_g^{(i-1)}]$. To study the convergent properties of the map we expand $g(\phi)$ near the fixed point ϕ^*, writing $\phi_g^{(i)} = \phi^* + \delta\phi_g^{(i)}$ [16]. This gives $\delta\phi_g^{(i+1)} = g'(\phi^*)\delta\phi_g^{(i)}$ so for a convergent orbit we require $|g'(\phi^*)| < 1$, and quick convergence can be expected for $g'(\phi^*) \approx 0$. We calculate $|g'(\phi)| - 1$ in the ϕ, χ space and observe that near the binodal, convergence is fast in the high-χ regime with $|g'(\phi)| \sim 0$ while much slower near criticality $\chi \sim \chi_c = 2$, and becomes 0 at exactly the critical point (figure 1D). We thus need the initial guess to be close to the binodal just at $\chi \sim \chi_c$ and both the spinodal and approximate binodal may seem to be appropriate choices. It is worth writing down these concentrations near criticality with $N = 1$: $\phi_\pm^{\text{spi}} \approx \frac{1}{2} \pm \sqrt{\frac{\delta\chi}{8}}$ and $\phi_\pm^{\text{bin}} \approx \frac{1}{2} \pm \sqrt{\frac{3\delta\chi}{8}}$. Furthermore, we can also obtain the approximate form of the $|g'(\phi)| - 1 = 0$ contour near $\chi_c = 2$ and $\phi_c = \frac{1}{2}$, which gives exactly $\delta\phi^2 = \frac{\delta\chi}{8}$. The spinodal thus coincides with the metastable line and is not an appropriate choice, despite it having better behaviour than the approximate binodal at large χ: the latter enters the unphysical regions $\phi < 0$ and $\phi > 1$ while the spinodal is always bound in $0 < \phi < 1$. We thus use the Ginzburg-Landau binodal (7) as the initial guess so

$$\phi_{\pm,g}^{(i)} = g^i\left[\frac{1}{2} \pm \sqrt{\frac{3\delta\chi}{8}}\right], \tag{9}$$

and the contraction mapping principle allows us to write the solution as

$$\phi_\pm^* = \lim_{i\to\infty} \phi_{\pm,g}^{(i)}. \tag{10}$$

Computing $\phi_g^{(i)}$ shows the 1D map [equation (8)] converges to the true binodal (figure 1E) in just 2 iterations. The convergence is however slow in the intermediate $\chi \sim 4$ range and the map can be improved by doing a self-consistent expansion.

Self-consistent expansion. Starting from the self-consistent equation $\phi = g(\phi)$ we approximate $g(\phi)$ as a linear function and solve the local self-consistent equation exactly. Writing $g(\phi + \delta\phi) \approx g(\phi) + g'(\phi)\delta\phi$ the self-consistent equation reads $\phi + \delta\phi = g(\phi) + g'(\phi)\delta\phi$, solving for $\delta\phi$ we obtain $\delta\phi = \frac{\phi - g(\phi)}{g'(\phi) - 1}$ so the new 1D map $\phi^{(i+1)} = \phi^{(i)} + \delta\phi^{(i)}$ becomes:

$$\phi_h^{(i+1)} \equiv \frac{g'[\phi^{(i)}]\phi^{(i)} - g[\phi^{(i)}]}{g'[\phi^{(i)}] - 1} \equiv h[\phi^{(i)}] \tag{11}$$

and it reduces to the original map when $g'[\phi^{(i)}] = 0$. The convergent property can be studied by expanding the above with $\phi_h^{(i)} = \phi^* + \delta\phi_h^{(i)}$, we arrive at

$$\delta\phi_h^{(i+1)} = \frac{g''(\phi^*)[g(\phi^*) - \phi^*]}{[g'(\phi^*) - 1]^2}\delta\phi_h^{(i)} \tag{12}$$

and near the fixed point the numerator approaches 0, so the convergence is rapid. Using the same Ginzburg-Landau expansion we have

$$\phi_{\pm,h}^{(i)} \equiv h^i \left[\frac{1}{2} \pm \sqrt{\frac{3\delta\chi}{8}}\right] \tag{13}$$

and the convergence is rapid (figure 1F).

Closed form expression. Here we propose a simple analytical form for the dilute phase binodal concentration. The dense phase results are obtained by simply reflecting the dilute phase concentration about $\phi_c = \frac{1}{2}$. We use a modified initial guess for the dilute phase concentration

$$\psi_-^{(0)} = \frac{1}{2}\left[\left(\frac{1}{2} - \sqrt{\frac{3\delta\chi}{8}}\right) + \left|\frac{1}{2} - \sqrt{\frac{3\delta\chi}{8}}\right|\right]. \tag{14}$$

This initial guess is bound within $0 \le \psi_-^{(0)} < 1$. Applying $g(\phi)$ once we obtain

$$\psi_-^{(1)} \equiv \begin{cases} \frac{1}{1+\exp\left(\chi\sqrt{\frac{3\delta\chi}{2}}\right)} & \text{if } \delta\chi \le \frac{2}{3} \\ \frac{1}{1+\exp(\chi)} & \text{if } \delta\chi > \frac{2}{3} \end{cases} \tag{15}$$

and the dense phase concentration ψ_+ is

$$\psi_+^{(1)} \equiv \begin{cases} \frac{1}{1+\exp\left(-\chi\sqrt{\frac{3\delta\chi}{2}}\right)} & \text{if } \delta\chi \le \frac{2}{3} \\ \frac{1}{1+\exp(-\chi)} & \text{if } \delta\chi > \frac{2}{3} \end{cases}. \tag{16}$$

These expressions cover both the low and high-χ regime (figure 1G). Near the critical point $\delta\chi \approx 0$ we have $\phi_2 \approx \frac{1}{2+\sqrt{6\delta\chi}} \approx \frac{1}{2} - \sqrt{\frac{3\delta\chi}{8}}$ as expected; on the other hand for large $\chi \gg 1$ we have

$$\psi_-^{(1)} \approx e^{-\chi}, \tag{17}$$

which is qualitatively different from the polynomial scaling of the spinodal concentration $\phi^{\text{spi}} \approx \frac{1}{2\chi}$. This exponential scaling arises from the chemical potential in the dilute limit taking the form of $\mu \approx \ln\phi$. An important implication is thus that the binodal phase separation can occur over a concentration range spanning orders of magnitude, while spinodal decomposition has a much narrower band of concentrations.

The proposed analytical form $\psi_\pm^{(1)}$ exhibits a discontinuity at $\delta\chi = \frac{2}{3}$ due to the discontinuity in the initial guess, and this can be smoothened out by introducing an additional parameter Δ in the following manner:

$$\tilde{\psi}_-^{(0)} = \frac{1}{2}\left[\left(\frac{1}{2} - \sqrt{\frac{3\delta\chi}{8}}\right) + \sqrt{\left(\frac{1}{2} - \sqrt{\frac{3\delta\chi}{8}}\right)^2 + \Delta^2}\right] \tag{18}$$

so that the first-order binodal is

$$\tilde{\psi}_-^{(1)} = \frac{1}{1 + \exp\left[-2\chi\tilde{\psi}_-^{(0)} + \chi\right]} \tag{19}$$

and the parameter Δ can be determined by numerically fitting $\tilde{\psi}_-^{(1)}$ to the exact binodal. This gives $\Delta = 0.185$ and provides a better approximate (figure 1H). As will be shown later, the same Δ value can be used in the case of general N with an appropriate scaling, and no further numerical tuning of Δ is needed.

IV. SELF-CONSISTENT SOLUTION FOR GENERAL N

Self-consistent equation. Here we extend the $N = 1$ solution to the general case. Denoting the polymer volume fractions in the dense and dilute phases as Φ_+ and Φ_-, the binodal equations are, using (5) with (6):

$$\begin{aligned} \left(\frac{\Phi_-}{\Phi_+}\right)^{1/N} \frac{1-\Phi_+}{1-\Phi_-} &= \exp[-2\chi(\Phi_+ - \Phi_-)] \\ \frac{1-\Phi_+}{1-\Phi_-} &= \exp\left[-\gamma(\Phi_+ - \Phi_-) - \chi(\Phi_+^2 - \Phi_-^2)\right] \end{aligned}. \tag{20}$$

Define the two exponentials $x \equiv \exp[-2\chi(\Phi_+ - \Phi_-)]$ and $y = \exp\left[-\gamma(\Phi_+ - \Phi_-) - \chi(\Phi_+^2 - \Phi_-^2)\right]$ the above is simplified to $\Phi_-/\Phi_+ = (x/y)^N$. Substituting this back into the second equation of (20) we get $\frac{1-\Phi_+}{1-(x/y)^N\Phi_+} = y$. Solving this in terms of Φ_+, and organising Φ_+, Φ_- in vector form:

$$\Phi \equiv \begin{pmatrix}\Phi_+ \\ \Phi_-\end{pmatrix} = \begin{bmatrix} \frac{1-y}{1-(x/y)^N y} \\ \frac{1-y}{(x/y)^{-N} - y} \end{bmatrix} \tag{21}$$

Define the operator on the right hand side of (21) as G we have the map

$$\Phi_{\pm,G}^{(i+1)} = G_\pm[\Phi^{(i)}]. \tag{22}$$

A scaling law for the dilute phase at large interaction strength can be obtained as before. At large χ we approximately have $\Phi_+ \approx 1$ and $\Phi_- \approx 0$, giving $x \approx e^{-2\chi}$ and $y \approx e^{-\gamma-\chi}$. The self-consistent expression for ϕ_- then becomes $\Phi_- \approx \frac{1-e^{-\gamma-\chi}}{e^{N(\chi-\gamma)} - e^{-\gamma-\chi}}$. This can be further approximated to be

$$\Phi_- \approx e^{-N(\chi-\gamma)} \tag{23}$$

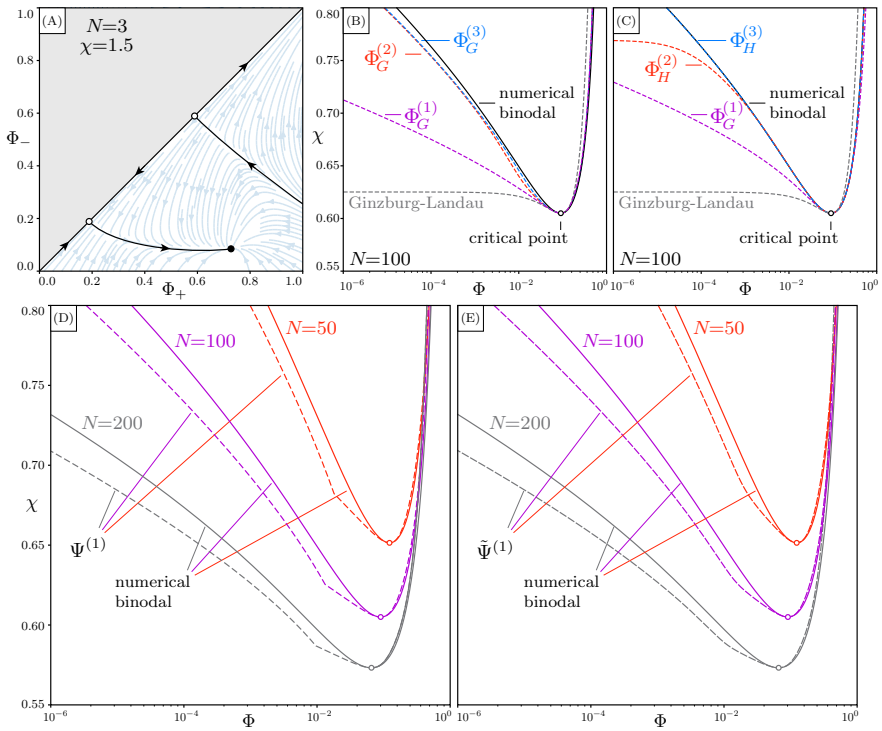

FIG. 2. Self-consistent solution for general N. (A) Flow field of the continuous map with $N = 3$, $\chi = 1.5$. Solid circles are stable fixed points corresponding to the binodal, and hollow circles are saddle points. (B), (C) Comparison between the numerical binodal (black solid line) with self-consistent schemes with 0, 1, 2 or 3 iterations (coloured dashed lines). $\Phi_G^{(i)}$ is the original map and $\Phi_H^{(i)}$ the improved map. The improved map is indistinguishable from the numerical results after 3 iterations. (D), (E) the proposed analytical forms $\Psi^{(1)}$, $\tilde{\Psi}^{(1)}$ are good approximations for binodal concentrations.

In the case of $N = 1$ we recover the $e^{-\chi}$ scaling. The behaviour of the 2D map can also be studied similar to the 1D case. Writing $\Phi_{\alpha,G}^{(i)} = \Phi_\alpha^* + \delta\Phi_{\alpha,G}^{(i)}$ with $\alpha = \pm$ we get $\delta\Phi_{\alpha,G}^{(i+1)} = [\partial_\beta G_\alpha(\Phi^*)]\delta\Phi_{\beta,G}^{(i)} \equiv J_{\alpha\beta}\delta\Phi_{\beta,G}^{(i)}$, with implicit summation of repeated indices. Stability requires eigenvalues λ_α of the Jacobian $J_{\alpha\beta}$ obey $|\lambda_\alpha| < 1$. Since both the eigenvalues and eigenvectors of $J_{\alpha\beta}$ are in general complex, to better visualise the convergence of the orbit $\Phi_G^{(i)}$ we instead promote the discrete map to a continuous flow equation parametrised by t: $\Phi_G \to \Phi_G(t)$, with $\dot{\Phi}_{\pm,G} = G_\pm(\Phi_G) - \Phi_{\pm,G}$. The velocity field $\dot{\Phi}_{\pm,G}$ then contains the behaviour of the orbit $\Phi_G^{(i)}$ in the limit of small time steps and four fixed points can be identified: 2 stable fixed points corresponding to the binodals and 2 unstable fixed points on the $\Phi_+ = \Phi_-$ diagonal (figure 2A). Using the Ginzburg-Landau binodal (7) as the initial guess we observe the orbit $\Phi_G^{(i)}$ is indeed convergent (figure 2B). The convergence is however still slow away from the critical point, and we use the self-consistent expansion again.

Self-consistent expansion. We adopt the linear approximation and write $\Phi_\alpha + \delta\Phi_\alpha = G_\alpha(\Phi) + J_{\alpha\beta}\delta\Phi_\beta$ and thus $\delta\Phi_\alpha = (J-1)^{-1}_{\alpha\beta}[\Phi_\beta - G_\beta(\Phi)]$. The improved map is

$$\Phi_{\alpha,H}^{(i+1)} = (J-1)^{-1}_{\alpha\beta}[J_{\beta\eta}\Phi_\eta - G_\beta(\Phi)]|_{\Phi=\Phi_H^{(i)}} \equiv H_\alpha[\Phi^{(i)}]. \tag{24}$$

Good agreement with numerical results is achieved within three iterations for a large $N = 100$ (figure 2C). The self-consistent solution is thus able to cover a wide range of N with very few iterations, and can be used to perform numerical fitting of experimental binodal data.

Closed form expression. To obtain good convergence within one iteration at large N we treat the dilute and dense phases separately. The dense phase binodal is close to the true binodal after one self-consistent step so for the dense branch we simply use $\Psi_+^{(1)} \equiv G_+\left(\phi_c \pm \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}\right)$, which gives, after simplification:

$$\Psi_+^{(1)} = \frac{1 - \exp\left[-\sqrt{\frac{6\delta\chi}{\chi_c^2\sqrt{N}}}\,(\gamma + 2\chi\phi_c)\right]}{1 - \exp\left\{-\sqrt{\frac{6\delta\chi}{\chi_c^2\sqrt{N}}}\,N\left[2\chi(1-\gamma\phi_c) - \gamma^2\right]\right\}} \tag{25}$$

where again we have $\gamma \equiv 1 - \frac{1}{N}$, $\phi_c \equiv \frac{1}{1+\sqrt{N}}$, $\chi_c \equiv \frac{1}{2}\left(1 + \frac{1}{\sqrt{N}}\right)^2$ and $\delta\chi \equiv \chi - \chi_c$. For the dilute phase concentration we use an approach similar to the $N =$

1 case. First we constrain the Ginzburg-Landau dilute phase concentration by writing

$$\Psi_-^{(0)} \equiv \frac{1}{2}\left[\left(\phi_c - \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}\right) + \left|\phi_c - \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}\right|\right]. \tag{26}$$

Applying $G(\Phi)$ to $\left[\Psi_+^{(0)}, \Psi_-^{(0)}\right]$ where $\Psi_+^{(0)} = \phi_c + \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}$, and using the dilute branch then gives

$$\Psi_-^{(1)} = \frac{1 - e^{-Q(\gamma+\chi P)}}{e^{-NQ[\gamma+\chi(P-2)]} - e^{-Q(\gamma+\chi P)}} \tag{27}$$

where $P \equiv \Psi_+^{(0)} + \Psi_-^{(0)}$ and $Q \equiv \Psi_+^{(0)} - \Psi_-^{(0)}$. Together $\Psi^{(1)} \equiv \left[\Psi_+^{(1)}, \Psi_-^{(1)}\right]$ gives a good approximation for the binodal concentrations at large N (figure 2D).

To smoothen out the kink we can use the Δ parameter again. We write the new dilute phase initial guess as

$$\tilde{\Psi}_-^{(0)} \equiv \frac{1}{2}\left[\left(\phi_c - \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}\right) + \sqrt{\left(\phi_c - \sqrt{\frac{3\delta\chi}{2\chi_c^2\sqrt{N}}}\right)^2 + \left(\frac{\Delta}{\sqrt{N}}\right)^2}\right] \tag{28}$$

with the original Δ scaled by a factor of N. The scaling mimics the scaling of $\phi_c \approx \frac{1}{\sqrt{N}}$ since Δ determines the concentration around which the initial guess transitions from the Ginzburg-Landau solution to the physical boundary. The new analytical form for the dilute phase concentration is then

$$\tilde{\Psi}_-^{(1)} = \frac{1 - e^{-\tilde{Q}(\gamma+\chi\tilde{P})}}{e^{-N\tilde{Q}[\gamma+\chi(\tilde{P}-2)]} - e^{-\tilde{Q}(\gamma+\chi\tilde{P})}} \tag{29}$$

where $\tilde{P} \equiv \Psi_+^{(0)} + \tilde{\Psi}_-^{(0)}$ and $\tilde{Q} \equiv \Psi_+^{(0)} - \tilde{\Psi}_-^{(0)}$. and this amendment removes the kink (figure 2E). For completeness we also define

$$\tilde{\Psi}_+^{(1)} = \Psi_+^{(1)} \tag{30}$$

so $\tilde{\Psi}_\pm^{(1)}$ gives the improved analytic form.

V. DIRECT DATA-FITTING

The self-consistent solution allows efficient computation of binodal concentrations and we use it to fit experimental LLPS data and extract the interaction parameters. In the following fitting 5 self-consistent steps according to equation (24) are used to compute the binodal concentrations. Binodals for the prion-like low-complexity domain from isoform A of human hnRNPA1 (A1-LCD) were measured in [17] (137 amino acid residues) and three series of A1-LCD variants are fitted here. The first series involves aromatic residues Tyrosine (Y) and Phenylalanine (F); the second series involves non-equivalent polar spacers Glycine (G) and Serine (S); and the third series involves ionic residues Aspartic acid (D), Arginine (R) and Lysine (K). The wild type (WT) A1-LCD binodal was also measured.

Fitting results. The chain length N is set as a global fitting parameter. The interaction parameter χ has the form $\chi = \frac{\Delta\epsilon}{k_B T}$ and we fit $\Delta\epsilon$ for each variant. To convert concentrations to volume fractions we use a protein density of 1.35g/cm^3 and average molecular weight of 13.1kDa to obtain the conversion ratio from concentration c(M) to volume fraction ϕ as $\phi = \frac{13.1}{1.35}\frac{c}{\mathrm{M}}$. The fitting results give the effective chain length $N = 182.7$, larger than 137, the number of residues. This is as expected as N represents the number of lattice sites occupied by each polymer, and since the lattice site volume is determined by the underlying medium, i.e. water, the effective N will be larger than the number of residues owning to the larger size of amino acids compared to water molecules. The ratio $\eta \equiv \frac{N}{137} = 1.33$ then represents the average number of lattice sites occupied by one residue. The fitted $\Delta\epsilon$ values represent the site-to-site contact energy and a larger $\Delta\epsilon$ indicates stronger attraction between proteins. To highlight the difference across variants we first calculate the protein-to-protein contact energy $E \equiv -N\Delta\epsilon$ and define the deviation from WT as $\Delta E_{\text{variant}} \equiv E_{\text{variant}} - E_{\text{WT}}$. ΔE results are listed in table I. Each variant series then allows quantitative interpretation of impacts of different residues on LLPS propensity.

Tyrosine is a stronger sticker than Phenylalanine. In the aromatic series the effective interaction energy agrees with the observation that Tyr residues are stronger stickers than Phe residues [17], and from the fitted value we can further infer the individual contribution of each Tyr and Phe residue. Denote individual residue contributions as ΔE_{Tyr} and ΔE_{Phe}, values from table I allow us to write down the following system of equations

$$\begin{cases} -12\Delta E_{\text{Phe}} + 12\Delta E_{\text{Tyr}} &= -6.3\text{kJ/mol} \\ +7\Delta E_{\text{Phe}} - 7\Delta E_{\text{Tyr}} &= +8.2\text{kJ/mol} \\ -4\Delta E_{\text{Phe}} - 2\Delta E_{\text{Tyr}} &= +23.8\text{kJ/mol} \end{cases} \tag{31}$$

which can be solved to give $\Delta E_{\text{Tyr}} \approx -4.1$ kJ/mol and $\Delta E_{\text{Phe}} \approx -3.7$ kJ/mol. Both Tyr and Phe are thus stickers with Tyr stronger than Phe.

Serine destabilises condensates. Performing calculations similar to above we extract the difference $\Delta E_{\text{Gly}} - \Delta E_{\text{Ser}} \approx -0.68$ kJ/mol, indicating destabilising effect of Serine residues. This can be understood as the OH group in Ser forming favourable interactions with water, thus destabilising the condensate.

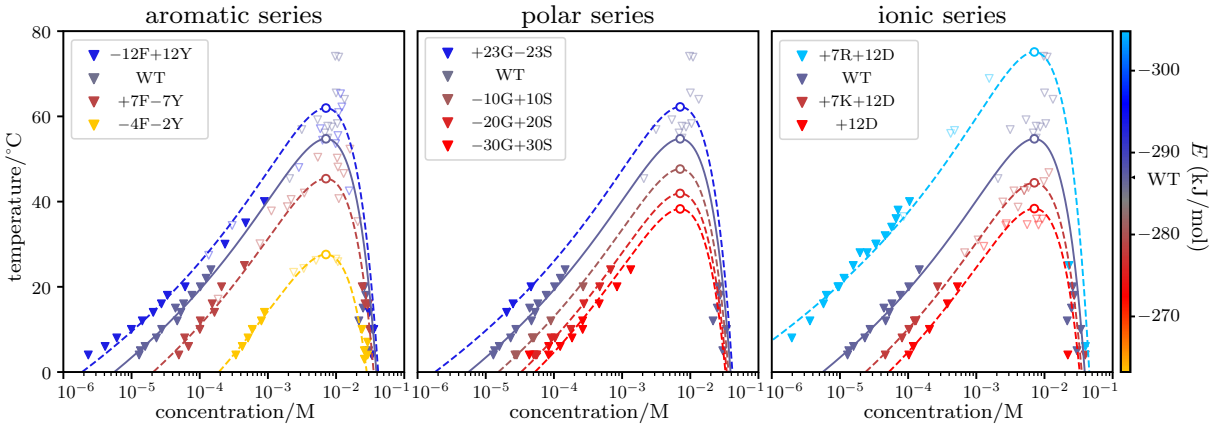

FIG. 3. Flory-Huggins fit of binodal data from [17], fitted using five self-consistent steps defined by equation (24). A constant N is maintained across all variants. Solid triangle markers are dilute and dense phase concentration measurements, and light hollow triangle markers are estimates of the critical point from cloud point measurements. Dashed lines are the best-fit curves and hollow circles are critical points. The WT binodal is the same in all three plots and is plotted in solid line. Colours of the plot represent the $E \equiv -N\Delta\epsilon$ values of the variant. The $\pm nX$ in variant names indicate n of X residues are added (+) or removed (−) from WT.

$N = 182.7 \quad E_{\mathrm{WT}} = -287.0$ kJ/mol

aromatic series		polar series		ionic series	
variant	ΔE (kJ/mol)	variant	ΔE (kJ/mol)	variant	ΔE (kJ/mol)
−12F+12Y	−6.3	+23G−23S	−6.6	+7R+12D	−17.8
+7F−7Y	+8.2	−10G+10S	+6.2	+7K+12D	+9.0
−4F−2Y	+23.8	−20G+20S	+11.2	+12D	+14.3
		−30G+30S	+14.4		

TABLE I. Fitting results for the A1-LCD data.

Arginine promotes LLPS better than Lysine at constant charge. The ionic series data is harder to interpret since the overall protein charge can have a non-monotonic effect on LLPS propensity [17]. We can however still compare the +7R+12D and +7K+12D variants since they have the same overall charge. The energy difference between Arginine and Lysine is $\Delta E_{\mathrm{Arg}} - \Delta E_{\mathrm{Lys}} \approx -3.8$ kJ/mol, indicating stronger sticker behaviour for Arginine. This can arise due to the electron delocalisation in the guanidinium group and higher charge-charge contact efficiency with other charged residues.

Fitting using analytic form gives similar insights. To test the applicability of equations (29) and (30) we fitted the same data to the analytic form. The fitting quality is similar to before (figure 4) while the parameters differ in value (table II). We note, however, that the N and E_{WT} values drift due to the gap between the analytic form and the true binodal, while the energy differences ΔE are still close to the previous fitting results. Practically, ΔE values are more important in understanding how individual amino acids impact LLPS propensity so the analytic form is still a valuable tool.

VI. CONCLUSION

We have developed a self-consistent solution for the binodal concentration of the 2-component Flory-Huggins phase-separating system. The proposed self-consistent operators shed light on the scaling behaviour of the dilute phase binodal, which is qualitatively different from the scaling of the spinodal and explains why LLPS of proteins occur over a concentration range spanning several orders of magnitude. Using the well-known Ginzburg-Landau binodal approximate solution as the initial guess, the self-consistent solution achieves numerical accuracy within 2 to 3 iterations and allows highly efficient fitting of experimental binodal data. Explicit analytical forms of the binodal concentrations are also proposed to approximate the binodal within one self-consistent iteration. Using the developed solution we fitted experimental data measured for variants of

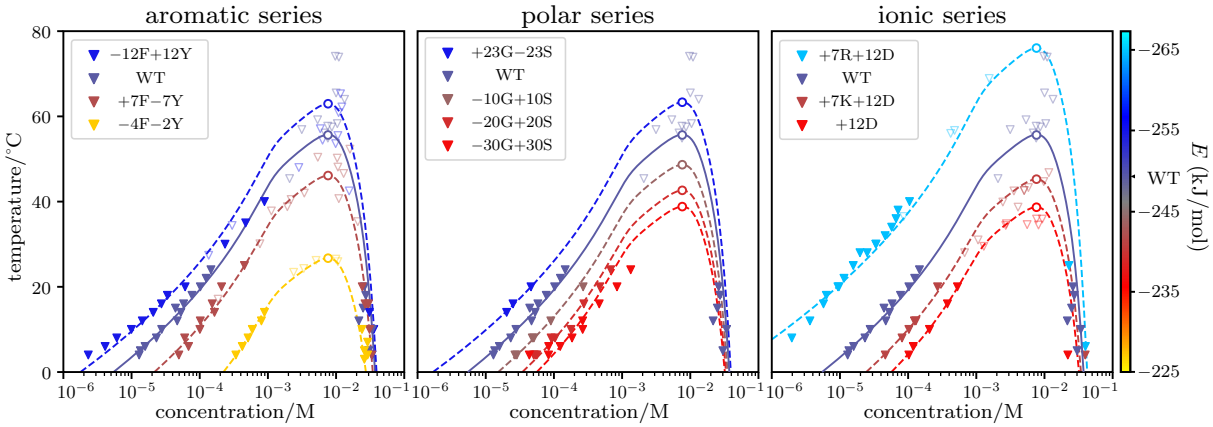

FIG. 4. Flory-Huggins fit of binodal data from [17], fitted using the analytic form [equations (29), (30)]. The quality of the fit is similar to the self-consistent steps.

$N = 156.5 \quad E_{\mathrm{WT}} = -249.3$ kJ/mol

aromatic series		polar series		ionic series	
variant	ΔE (kJ/mol)	variant	ΔE (kJ/mol)	variant	ΔE (kJ/mol)
−12F+12Y	−5.5	+23G−23S	−5.7	+7R+12D	−15.3
+7F−7Y	+7.3	−10G+10S	+5.4	+7K+12D	+7.9
−4F−2Y	+22.0	−20G+20S	+9.9	+12D	+12.9
		−30G+30S	+12.8		

TABLE II. Fitting results for the A1-LCD data with the analytic form. The N and E_{WT} values are slightly different from the results obtained using the self-consistent steps but ΔE across variants are similar to previous fitting results.

the A1-LCD protein and extracted effective interaction energies, which can be used to further decode the impact of individual amino acid residues on LLPS. Our analytical solution to the Flory-Huggins model thus allows systematic investigation of sequence grammar of LLPS-prone proteins and can lead to development of a wholistic framework for predicting LLPS propensity from sequence information.

Code availability. A python library containing the self-consistent maps and auxiliary functions can be found at: https://github.com/GrapeKing/Flory-Huggins.

Acknowledgement. This study is supported by the Laboratory for Molecular Cell Biology, University College London (T.C.T.M.) and the Wellcome Trust Collaborative Award 203249/Z/16/Z (T.P.J.K.). The authors thank Alexander Röntgen for helpful discussions.

Conflict of interests. The authors declare no conflict of interests.

[1] G. L. Dignon, R. B. Best, and J. Mittal, Biomolecular phase separation: From molecular driving forces to macroscopic properties, Annual Review of Physical Chemistry **71**, 53 (2020).
[2] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry, Nature Reviews Molecular Cell Biology **18**, 285 (2017).
[3] Y. Shin and C. P. Brangwynne, Liquid phase condensation in cell physiology and disease, Science **357**, 10.1126/science.aaf4382 (2017).
[4] W. M. Babinchak and W. K. Surewicz, Liquid–Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation, Journal of Molecular Biology **432**, 1910 (2020).
[5] J. Söding, D. Zwicker, S. Sohrabi-Jahromi, M. Boehning, and J. Kirschbaum, Mechanisms for Active Regulation of Biomolecular Condensates, Trends in Cell Biology **30**, 4 (2020).
[6] J. M. Choi, A. S. Holehouse, and R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions, Annual Review of Biophysics **49**,

107 (2020).

[7] C. P. Brangwynne, P. Tompa, and R. V. Pappu, Polymer physics of intracellular phase transitions, Nature Physics **11**, 899 (2015).

[8] P. J. Flory, Thermodynamics of High Polymer Solutions, The Journal of Chemical Physics **10**, 51 (1942).

[9] P.-G. De Gennes and P.-G. Gennes, *Scaling concepts in polymer physics* (Cornell university press, 1979).

[10] S. Mao, D. Kuldinow, M. P. Haataja, and A. Košmrlj, Phase behavior and morphology of multicomponent liquid mixtures, Soft Matter **15**, 1297 (2019), arXiv:1810.03689.

[11] Y.-H. Lin, J. Wessén, T. Pal, S. Das, and H. S. Chan, Numerical Techniques for Applications of Analytical Theories to Sequence-Dependent Phase Separations of Intrinsically Disordered Proteins, arXiv , 1 (2022), arXiv:2201.01920.

[12] A. Granas and J. Dugundji, *Fixed Point Theory*, Springer Monographs in Mathematics (Springer New York, New York, NY, 2003).

[13] E. Zeidler, *Nonlinear Functional Analysis and its Applications: Part 1: Fixed-Point Theorems* (Springer New York, 1985).

[14] S. I. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. Knowles, Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments, Journal of Chemical Physics **135**, 10.1063/1.3608916 (2011).

[15] S. I. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. Knowles, Nucleated polymerization with secondary pathways. II. Determination of self-consistent solutions to growth processes described by non-linear master equations, Journal of Chemical Physics **135**, 10.1063/1.3608917 (2011).

[16] S. H. Strogatz, *Nonlinear Dynamics and Chaos* (CRC Press, 2018).

[17] A. Bremer, M. Farag, W. M. Borcherds, I. Peran, E. W. Martin, R. V. Pappu, and T. Mittag, Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains, Nature Chemistry 10.1038/s41557-021-00840-w (2021).